DOCUMENT RESUME

ED 185 864 FL 011 406

AUTHOR Tharu, J.
TITLE Rationalizing the Pass-Fail Decision: A First Step in Examination Reform.
INSTITUTION Central Inst. of English and Foreign Languages, Hyderbad (India).
PUB DATE 73
NOTE 19p.
JOURNAL CIT CIEFL Bulletin: v9 p41-58 1972-1973

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Achievement Tests; *English (Second Language); Essay Tests; *Language Tests; Minimum Competency Testing; Secondary Education; Standardized Tests; *Test Construction; Test Reliability; *Test Validity

ABSTRACT

The soundness of the existing system of examination at the high school level in India is assessed with particular reference to the English section. The purposes to be served by the examination are discussed and the shortcomings of the system are described. The test and the passing procedure is found inferior to those systems which utilize standardized tests. The fixed passmark and pass-fail decision are liable to serious error. The value of the test is pre-ordained by administrative sanction rather than by empiricial evidence. A scheme for making the pass-fail decision more systematic is suggested. In the new model, minimum requirements for passing are tested separately from those designed to order a candidate's achievement. Questions in the two-stage model differ substantially in difficulty. The assumptions about the hierarchical structure of language skills that underlie the model are similar to those that inform decisions about grading and presentation of instruction. Sample portions of the new model and discussion about the grading of this test are included. (PMJ)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Director, CIEFL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED185864

*CIEFL Bulletion*
9 (1972–73) 41–58

# Rationalizing the Pass–Fail Decision: A First Step in Examination Reform

J. THARU

## I. Approaches to test Validity

The classical defination of *validity* of a test refers to the extent to which it actually succeeds in measuring what it purports to measure. A number of specific aspects of the more general concept of validity such as content validity, concurrent validity, are widely used. The validity of measures of educational achievement has been seen largely in terms of content validity. The proper coverage of all parts of the relevant syllabus is clearly the major requirement for a sound test or examination. About ten years ago the clarification of instructional objectives, as discussed by Bloom (1956) and Mager (1962), became a central concern in testing. Since then a persistent emphasis on testing comprehension rather than mere memory, has featured in the conception of validity.

The major approaches to validity noted above, are all derived from the assumption that there exists some "true level" of ability or mastery of the subject matter. The problem in constructing a test is seen as one of attaining maximum accuracy in measuring this true value. This focus may be regarded as related to the inherent or structural validity of the test. A somewhat different, but complementary, approach to validity, particularly for educational tests, has emerged in recent years. This emphasizes the context in which a testing programme is initiated, and the use which is made of the results. It is not just accurate scores as such that make for good validity, but also a sound basis for their practical use. Carroll (1961) suggests that the purpose of testing is always to render information to aid in making intelligent decision about possible courses of action. Cooper (1972: 332) reports that, "all too often tests are administered continually, the original purpose of testing forgotten and the scores serving no purpose....." He insists that testing is not justified unless "decisions are made on the basis of test scores and unless the information provided by the test scores is relevant to these decisions". Cronbach (1970:89) has remarked that "since virtually no test is confined to a single purpose, it is illogical to speak of test validity. What one has to validate is a proposed interpretation of the test."

FL011406

This consideration—the soundness of the practical use made of the results suggests a basis for further studying the validity of examinations as we use them. Educational achievement tests can be put into one of two major categories: teacher made or informal tests, and standardized tests. The public examinations held at various level in our system are intended to serve more

or less the same purposes as standardized tests do, in a country like the U.S.A. During the past several decades of progress in the field of educational testing, a framework of standards and guidelines for judging the soundness and quality of tests has emerged. Most of the sophistication and knowhow has developed from the experience gained in constructing and widely using standardized achievement tests. Any attempt to evaluate examinations using this modern psychometric framework must, therefore, identify the significant points of difference between such *public examinations and standardized tests*. A discussion of steps to improve examinations must take due note of their unique characteristics, as much of the new 'technology' of psychometrics cannot be applied directly. In the discussion which follows, the English Language or general English paper at the High School/P.U.C. level will be taken as an example of the *examination paper* for detailed analysis. The generalizability of many observations made in this context to examinations of all subjects and levels will be readily apparent. However this article will restrict itself to a detailed consideration of the English examination paper, and not directly take up the wider issues involved.

## II. A comparison of the examination with the standardized test

The main features of the typical examination paper in English may be listed as follows.

(1) It carries exactly 100 marks and has a time allowance of 2½ to 3 hours.

(2) A number of different types of items are used: long essay, short paragraphs, short answers, and various forms of objective type items.

(3) The weightage (marks) given to each type of item may vary from 25 or 30 (for an essay) to as little as 2 or 1 (for a "supply the correct verb form" item).

(4) The selection of the items themselves and the allocation of weightage are based almost entirely on a priori and logical considerations rather than on empirical data from item analysis. A practical tryout is *not* one of the routine steps in the preparation of the final revision.

(5) The result is declared in terms of the raw scores or marks on the scale of 0-100. A passmark (assumed here to be 35) and other critical values are fixed beforehand.

(6) The decisions involved in (5) are a matter of statutory procedure, and are not reviewed in the light of empirical test statistics, pertaining to central tendency and dispersion. (The award of grace marks is not relevant to this point. See para 4 page 45.)

In comparison with this, a standardized test of achievement is markedly different in various ways. Firstly, the raw scores are never used to report the results; they are often never disclosed. Raw scores are very clearly regarded only as an intermediate step in the computing of derived scores, such as:

(i) *Percentile ranks,* which indicate where a student stands in a rank ordering of the group he is being compared with. Statements such as "just within the top 10%" or "ahead of the bottom 35%" are thus possible.

(ii) *Standard Scores,* which indicate the deviation of the students raw score from an empirically determined mean value (using that or an equivalent test) for the reference group.

Secondly, the items are almost entirely of the objective type. The test contains a large number of short items carrying the same weightage of 1 point. Occasionally 2 point items are also used. 3 point or 5 point items are virtually never used. The maximum raw score is thus determined largely by the number of items: it may be 275 or 475. There is no attempt to make it a convenient number like 100 or 500. Finally, though, it will not be apparent from just looking at the finished test, this final version is the end result of an essentially practical and empirical procedure of improving and refining a preliminary draft. The face value of items is important only in the preliminary form. Inclusion, with or without modification, in the final version is largely dependent on indications of suitability from the tryout results.

From the above brief comparative survey the examination paper stands out as a highly peculiar testing mechanism — with most of its value preordained by administrative sanction rather than by empirical evidence. In using raw scores on a scale of 100 and emphasizing the *per cent* fractional value, there is an underlying implication that the examination is measuring some ability with a potential maximum of "100 units". It is only such a premise that allows a prior decision that a student getting only 34 per cent marks shall fail. When, today, the use even of an interval scale in psychological measurement is a matter of disagreement, this unembarassed use of a ratio scale is questionable, to put it mildly. The practice of adding marks in various papers directly to get an aggregate (with no allowance for differences in range and variability of the separate distributions) adds to the inconsistency and irrationality of the system as a whole. It is worth noting that despite several years of keen interest in examination reform, and of research into examination procedures, no serious attempt has been made to put the whole body of practices into a comprehensive and consistent psychometric model or framework. Presumably this is a logically impossible task. However, it needs to be emphasized that without the development of an overall framework the multiplicity of piecemeal suggestions for reform — in themselves quite sound and pertinent —cannot have any worthwhile impact.

## III. The purposes to be served by the examination

The focus on validity which stresses the use (and usefulness) of test scores for making practical decisions has been noted earlier. This suggests a framework for an overall analysis of the present system which might in turn lead to the formulation of some desirable modifications. What are the critical decisions which are based on the examination score? Given these "fixed points" how best can the test making technology available today be utilized to facilitate and rationalize such decisions?

The major institutional or practical decision taken on the basis of examination marks is the declaration of *pass* or *fail,* with regard to each candidate. Certain second order decisions — the award of grades or divisions within the pass category — are also made. Since these latter steps are linked with the policy of adding marks from several papers to get an aggregate, they will not be studied in detail. The discussion will focus on the accuracy of the pass–fail decisions in the case of the English language paper. It is contended, however, that the basic conditions needed to rationalize this step, will also help to systematize the whole procedure for awarding marks.

Given the need to declare a student as having passed or failed, and working backwards as it were, the problem may be stated thus: to prepare a paper of 100 marks such that a student who did not master certain minimum essentials prescribed in the syllabus, will score 34 marks or less, while those who have met this standard will get 35 marks at least (and more in proportion to their further mastery, beyond this minimum level.)

It is an unfortunate fact that a search for "minimum requirements" for passing in syllabus statements will lead to the discovery that it is in fact "35 marks". This circularity of definition does not clarify in any way what is meant by passing. The assessment of the degree to which students "have mastered the minimum essentials of the course" was identified by Furst (1958: 192) as one of the distinct purposes of the achievement test. This and other insights contributed by Furst relating to planning a test have been lost sight of in recent years – perhaps because of an excessive concern with the type of item to be used rather than with the *objective* to be accomplished thereby. The need for a statement or description, in operational terms, of the level of language using ability regarded as the minimum acceptable, is quite obvious. While an actual statement cannot be developed here, a framework for doing so may be suggested.

The essential meaning of failure in an achievement test is the existence of a *need for the student to undergo the course of study again*–in its entirety or in a condensed (tutorial college) version. To say that a student has passed is to indicate that he can proceed to the next higher level of instruction in English and successfully meet the initial capability standard assumed there. Or, in the example of the PUC examination, if the student's formal education stops there, passing in English implies that he can cope with whatever English he may encounter in a trade or profession he enters with such a qualification. These considerations provide only a framework, and much more detailed specification of the actual performance eligible for a pass is necessary. The purpose here is only to emphasize the need for such a definition as part of a syllabus statement. This would yield an essential building block for the test -a conception of the level of functioning of the student who just "deserves to fail." This would be operational and understandable in terms of the domain of intended learning, and also allow this student to be distinguished from another hypothetical student who just "deserves to pass". When such operational definitions describing what is above and what is below the pass-fail transition point are available, some suitable arrangement of items and marks can be attempted. The need is for a procedure that will clearly and consistently differentiate in terms of marks between students just above and just

below the pass-fail point. In this context it is worth recalling a forceful plea made over a decade ago by George (1962: 58). He stressed the necessity for deciding "what characteristics are going to determine success and failure". Once such a decision has been made these critical points should appear in the examination, "not once but several times; so that each candidate fails not once but several times, or succeeds not once, but several times". The importance of this seemingly complicated procedure will be more readily seen, when the working of the present system is studied more carefully.

## IV. Shortcomings of the present system

The existing structure of the question paper regards every item as suitable for discriminating between *borderline* pass or fail students. Accordingly all the marks in all the questions are added up and the total so obtained compared with 35. What actually happens in a typical borderline case is something like this:

The student will get a few marks for each question. Some vocabulary items and some short grammar items will be successfully attempted by him. He may get anything between 5 and 10 marks for each of these sections assuming they carry about 20 marks. The comprehension questions and composition question (s) will also yield him about a third or more of the full marks alloted to them. His total is thus made up of these twos and threes and fives and sixes, and it may be 30 or 35 or 40. There is no clear and substantive basis for distinguishing between a student scoring, say, 38 marks and one scoring 31, which is consistently indicated by each separate question.

A further complication is added by the reluctance of examiners of essay type answers to use fully the range of marks available so that even a very poor essay often gets about 5 marks out of 20. This particular problem is discussed more fully by Grieve (1964:45). In statistical terminology the margin of error in this zone near 35 marks is very high. Such error is also found at all other levels, but this is not of relevance here. A recent report of the University Grants Commission (1973:11) states with great candour that "when a candidate is awarded a 'raw' mark of 41, the 'true' mark may be either above 46 or below 36 in fifty percent of the cases. Under these circumstances, the 10' point scale where candidates are distinguished in steps of one mark loses all significance." It is true that grace marks are often awarded However, the formula tends to take into account all considerations except psychometric ones, like the actual difficulty level of the paper. While this serves some humanitarian (or, perhaps, peace keeping) ends, it certainly does not make more accurate or meaningful a set of marks between about 30 and 40, which are virtually a set of random numbers.

Thus the figures on which the pass-fail decision depends now are altogether without any reliability or clear meaning. For students who are so obviously good or bad that a pass-fail decision can be easily made, the examination marks also give dependable and unambigious indications. However, it is in deciding the borderline cases, where accuracy is most needed and which is the main function of the examination, that we find the error margin to be so damagingly high. It needs to be stressed that the declaration of pass or fail

is a crucial academic-educational decision—one that must be fully within the control of the teacher. The term 'teacher' as used here is meant to include all subject specialists who might participate in the *planning* and implementation of instruction — of the syllabus, the curriculum materials, the examinations. It is meant to exclude those who (irrespective of academic status) can perform *only* a clerical function, when adjusting or moderating a table of marks *already awarded* by competent examiners.

Errors in the award of a distinction of a first division may be of personal significance to the individual concerned, but such an award is not, in itself, of much pedagogical significance. The pass-fail decision is fundamental to the whole evaluation process, and improving its dependability is an urgent necessity. Some procedure specifically designed to tackle this problem is needed. The several "improvements" in paper setting — closer coverage of syllabus, clear farming of questions, definite guidelines for marking, etc. —all of which are being seriously attempted now, do not materially alter the error margin at the pass-fail boundary. Indeed it is quite likely that all the advantages that go with the above steps are virtually wiped out by the inaccuracy of the pass-fail decision.

## V. An underlying principle of equivalence

The undependability of the marks obtained under the present system has two aspects. Firstly, their unreliability in relation to the stability demanded by a fixed cut-off point; and secondly, their meaninglessness even if some, reliability existed. This is the outcome of the procedure of adding up all marks from all questions indiscriminately, to get the simple total. The main thesis of this discussion is that an erroneous and untenable assumption relating to *the functional equivalence of equal units of numerical marks* underlies the existing system. Any model which overcomes the difficulties noted above, must begin by discarding this equivalence principle. The operation of this principle will be illustrated through the following table of marks in a 100 mark paper, for six students.

| Student | Maximum | A | B | C | D | E | F |
|---|---|---|---|---|---|---|---|
| *Section* | | | | | | | |
| 1. Essay | 30 | 10 | 10 | 14 | 15 | 15 | 22 |
| 2. Comprehension | 20 | 7 | 7 | 11 | 10 | 10 | 10 |
| 3. Vocabulary | 20 | 6 | 10 | 6 | 8 | 15 | 8 |
| 4. Short grammar items | 15 | 4 | 6 | 4 | 9 | 5 | 5 |
| 5. Miscellaneous objective type | 15 | 4 | 6 | 4 | 8 | 5 | 5 |
| Total | 100 | 31 | 39 | 39 | 50 | 50 | 50 |

*Comments:*

(i) Student A is a clear failure; B and C have passed; D, E, F are satisfactory.

(ii) Comparing A, B, and C we find that the difference between passing and failing is a matter of 8 marks. B manages this (in relation to A) by a few more correct words and a few short items. C writes a noticeably better essay, and also moves well ahead in the comprehension tasks. Both these separate forms of improvement, despite significant qualitative differences, are valued at 8 marks, making B and C "equally proficient".

(iii) Students D, E, F represent three equally good or equally mediocre, students. Here again, it is seen that a clearly superior composition written by F can be "balanced" by seven vocabulary words in the case of E, and a few short items in the case of D. These relative jumps of seven marks represent vastly different orders and levels of language ability but an arbitrary numerical equality is all that the marking scheme makes available.

The above example, using some actual figures has indicated two aspects of the principle of equivalence that seems to operate. One is that equal marks anywhere in the range of marks for a single question (like an essay) are taken at their face value. The "extra" capability that can improve an essay from 10/30 to 15/30 is seen as equal to that which can take an essay from 18/30 to 22/30. Secondly different areas of ability are simply equated. Thus coping with a number of comprehension tasks can be balanced by a list of vocabulary words. This second aspect is, to some extent, a necessary compromise that must be made when a complex array of skills is assessed by a single instrument. However, the arbitrary equating of skills and abilities at different levels of a hierarchical structure, is a totally unacceptable feature.

The new model presented below attempts to overcome these difficulties by making two significant departures from the conventional scheme of awarding marks. It imposes some restrictions on the equivalence principle. It also involves an indirect method of marking, so that all the raw scores or "points" do not feature as they are in the final tally of marks.

## VI. The proposed new structure for the examination paper

The model is based on a number of premises. Many of these principles have been mentioned in the preceding discussion. They are listed below in groups under appropriate sub-headings:

1. Concerning the subject matter or test domain:

(*a*) The course content of any language examination (except at a very elementary level) includes both simple and complex language abilities as learning targets.

(*b*) The simple skills are acquired earlier, and are to a large extent prerequisites for successful progress towards mastery of the more complex skills.

(*c*) The simpler abilities are either present or not. No intermediate condition can be specified which is of relevance to an achievement test. (This does not exclude its potential significance in a diagnostic context).

(*d*) The attainment of the more complex abilities is a matter of degree. A student may be at one of several stages of partial acquisition, at a given time.

2. Concerning the population of examinees:

(*a*) Any typical group of students taking the examination will represent a wide range of attainment of the objectives of the course. (This is a merely statistical statement: no postulate about differences in innate intellectual ability or in language learning aptitude is involved.)

(*b*) The weaker students (those who should properly fail) will show a mastery of some of the simple skills, but for the most part only a superficial familiarity with, rather than any worthwhile control of the operations constituting the complex skills.

(*c*) The mediocre and superior students will have a fairly complete mastery of the basics, and show varying degrees of tangible progress towards mastery of the complex skills.

3. Concerning the test items to be used:

(*a*) An item (question) used to assess *mastery* of a simple skill must elicit the appropriate behaviour in a number of situations, and demand a highly consistent pattern of success.

(*b*) An item to assess the degree of mastery of a complex or advanced skill should be at a noticeably higher level of difficulty (than items in 3 (a) above).

(*c*) Whenever possible separate items should be used for each aspect of a complex skill that can be treated in isolation.

4. Specifications for the test:

Taking into account the range of complexity in the domain, and the range of ability among the examinees, the following features are considered essential for a suitable test. These points represent the significant departures from the conventional pattern that the proposed model entails.

(*a*) The decision whether to pass or fail a student is made on the basis of a set of questions (called Stage I) designed specifically to test mastery of the points considered as minimum essentials.

(*b*) The result of stage I is a simple declaration of pass or fail. All those passing get 40 marks and all those failing get 25 marks. While these *exact* figures are arbitrary, the principle that each sub-group is not further subdivided is basic. A somewhat similar suggestion, regarding the creation of a bimodal distribution of examination marks, was made by George (1961:60).

(*c*) The finer gradation of students within the pass category is accomplished by another set of questions (Stage II), which discriminate among degrees of mastery of the advanced skills. This section carries 60 marks. In practice this section will be similar to a well constructed conventional paper, except for the absence of "easy" questions.

(*d*) The final award out of 100 marks, for passing students is (40+marks for Stage II). Students who have not passed Stage I should not in theory even attempt Stage II. In the absence of any suitable mechanism for such exclusion, at present, the procedure will be to ignore the marks (if any) gained in Stage II.

In addition to the above somewhat formally stated considerations, two central concerns that underly this approach to the examination paper, are worth identifying. Firstly, the critical pass-fail decision needs to be protected from the dependence on stray marks from a veritable jumble of items and item-types that is characteristic of the present set up. The questions of stage I are, therefore, designed specifically to indicate reliably, and with some degree of meaningfulness whether a candidate has attained a minimum standard or not. Secondly, the two stage arrangement seeks to avoid the extreme illogicality of asking a student who cannot even use pronouns correctly in a simple sentence, to "write a passage of 200 words" or "express in your own words the central theme..."(underlying 300 words of expository prose.) The illogicality lies not so much in the asking—but rather in seeking to award a specific mark of 3 or 4 or 5 (out of 20) to the resulting answer. Unless a minimum level of overall proficiency — consisting of a mastery of basic skills- is available the complex task cannot be meaningfully *approached,* let alone successfully undertaken.

An analogy from the psychomotor domain will help to clarify this argument. A novice tennis player needs first to learn how to hold the racquet firmly, to strike a slowly moving ball in a desired direction, etc. It is only then that it is relevant to ask him to face an opponent with a powerful service, who imparts speed and spin to the ball and causes it to bounce at a low angle from some difficult corner. The response of a suitably skilled player (which may or not actually win that point) would include judging the path of the ball after it bounces: moving to a suitable position, striking the ball in a tactically sound direction, so that the opponent gets no advantage from the return shot, etc. If the novice is put in this situation he will presumably run towards the ball, hit out with his racquet and perhaps even send the ball back. All these awkward movements are undoubtedly steps in the right direction, but never is there any serious attempt to assess a beginner's progress through such observations. Tne novice player is simply not put into such irrelevant situa-

tions. Similarly it is argued that some candidates cannot meaningfully be asked to attempt the more advanced questions.

As noted earlier, no proper specification of minimum requirements for passing is available for any level. For the purposes of illustrating the type of examination paper under discussion, some *provisional* statement is necessary. Taking up a high school level English Course (as represented by Book V of the Special Series Readers*) the following general declaration of minimum requirements is made.

Firstly, the student should have an active control over the vocabulary, sentence structures function words, that occur in the Reader, which are not separately listed as "new", and hence may be presumed to have been presented in earlier years. Secondly, and by the same token, he should have a recognition level knowledge of those new items which constitute the main teaching points of the lessons. These criteria will determine the level of difficulty aimed at in the questions of Stage I. It follows that Stage II would cover essentially the same material, but require a higher degree of mastery and control.

It needs to be emphasised that the above specification is solely for the purposes of illustrating the questions to be used. It is not an essential feature of the model. Sample questions for each stage are now presented.

---

*Stage I*: Answer all the questions.
All the parts of each question must be answered.

Q. 1. Fill in the blanks with a suitable personal pronoun : I, me, my, etc.

(*a*) I gave————ticket to the Inspector.

(*b*) They will collect———wages on Saturday.

(*c*) Please bring————friends to the wedding.

(*d*) He saw ———at the cinema on Monday.

(*e*) She asked————to bring this parcel here.

(*f*) I want to talk to————just now.

(*g*) The Principal told the students that————request for a holiday would be granted.

(*h*) ————used to go to the park every Sunday morning.

* Prepared by the CIEFL and published by the NCERT (1971).

Q. 2. Fill in the blanks with a suitable word from the following : who, where, when, which.

(*a*) This is not the bus on ————we will travel.

(*b*) The police want the person———saw the accident.

(*c*) This is the place ————the accident took place.

(*d*) The prisoner may be questioned only————his lawyer is present.

(*e*) The match will begin ———the Chief Guest has taken his seat.

(*f*) The express train will not stop at the place———we want to get off.

(*g*) These are districts————there is heavy rainfall for nearly half the year.

(*h*) The Minister mentioned several factors————made it unsafe to build the new dam there.

Q. 3. Form a complete sentence with each list of words. Add *the, an, a, some* where necessary, and supply proper punctuation marks.

(*a*) pencil table inkbottle on near is
(*b*) blue bag my has who
(*c*) dogs barking small are loudly
(*d*) postman brother my letter to gave
(*e*) child crying is now why
(*f*) cinema who will us tickets for buy
(*g*) just Ashok joined school has
(*h*) book as well untruthful dull as is

Q. 4. Each part of this question contains a sentence, which is followed by two statements, A and B. Indicate which statement, A or B, gives the correct meaning of the sentence.

(*a*) The more money he gets the more he wants.

A. He wants more money and he is able to get it.
B. He has got plenty of money, but he still wants more.

(*b*) They should have built their house far from the road.

A. Their house is not far from the road and this is a bad thing.
B. Their house is far from the road and this is a good thing.

(*c*) You can use my pen as long as you write carefully,

A. If you will write carefully, you can use my pen.
B. If you want to use my pen for a long time you must write carefully.

(*d*) But for his help we would not have reached the station on time.

A. Since he did not help we did not reach the station on time.
B. It is only because he helped us that we reached on time.

(*e*) Keep some candles with you in case the electricity is cut off.

A. The electricity will be cut off and you will need the candles then.
B. The electricity may be cut off and you may need the candles.

(*f*) Now that the rain has stopped the tennis match can continue.

A. The match can continue only when the rain stops.
B. The rain has stopped, and it is possible to continue the match.

Q. 5. Fill in the blanks with the most suitable word.

(*a*) He has lived in India for many years and is————-with our customs.
peculiar habitual familiar cooperative

(*b*) The heavy flooding of the roads was his————————for coming late.
promise excuse complaint attempt

(*c*) The young girl who won the quiz prize showed a ———————— understanding of politics.
remarkable breakable regular honest

(*d*) The success of this irrigation project depends on how much the farmers————————.
cooperate complete communicate complicate

(*e*) The old jeep is used ———for bringing water from the stream to the house now.
formerly really mainly extremely

(*f*) The lady was trying to————to the guard how she lost her box on the train.
obtain retain ascertain explain

(*g*) During the monsoon months it is difficult to maintain———— with the people on the island.
communication community confirmation competition

(*h*) The examinations will ————on March 15th this year.
complete proceed postpone commence

(*i*) Most Government offices————old records and files for a very long time.
reserve request acquire preserve

(*j*) He developed the————of rising early while he was at school.
habit imagination profit cultivation

Q. 6. Make sentences which will show clearly the meaning of each of the following words.

(*a*) remember
(*b*) expenses
(*c*) lend
(*d*) weaker
(*e*) special
(*f*) repeat
(*g*) wealthy
(*h*) distance

*Notes on the marking scheme for Stage I* :

(1) Each question has a large number of parts, each of which carries a *point*. A student getting 75% or more of the points will get the full *marks* for the question. Otherwise he will get no marks at all.

(2) Assuming that where are 10 questions, each carrying 4 marks, a student may get 40, 36, 32, .....(some multiple of 4) . If the pass mark is set at 30, a student "clearing" eight or more of the ten questions passes while all others fail. All those who pass get 40 marks, while all those who fail get 25 marks.

*Stage II* :

Q. 1. Choose the word or phrase nearest in meaning to the problem word :

(*a*) magistrate :
policeman judge witness lawyer

(*b*) reconcile :
bring together study in detail
rewrite in proper language enlarge

(*c*) petition :
request command regulation explanation

(*d*) reputation :

what people like about one fame
what people think about one popularity

(*e*) Seize :

stop convert connect take

(*f*) convenient :

bearable suitable religious expensive

(*g*) essential :

accessory visible necessary valuable

(*h*) undergo :

bury suffer weaken remember

(*i*) isolate :

separate from others prominent in a group
become similar following after

(*j*) confident :

knowing the answer expecting to succeed
holding a high post supporting the right side

Q. 2. Write a sentence to illustrate the meaning of each of the following words or phrases :

(*a*) in spite of (*b*) once and for all
(*c*) in favour of (*d*) do justice to
(*e*) characteristic of (*f*) clear (adjective)
(*g*) clear (verb) (*h*) clearly
(*i*) clearing (noun) (*j*) clarity

Q. 3. (1) Read the following passage carefully :

As a student I had heard that the lawyer's profession was a liar's profession. But this did not influence me, since I had no intention of earning either position or money by lying.

My principle was put to the test many times in South Africa. Often I knew that my opponents had coached their witnesses ; and if I only encouraged my client or his witnesses to lie, we could win the case. But I always resisted the temptation. I remember only one occasion when, after having won the case, I suspected that my client had deceived me. In my heart of hearts I always wished that I should win only if my client's case was right. In fixing fees, I do not recall ever having made them conditional on my winning the case. Whether my client won or lost, I expected nothing more nor less than my fees.

Now answer all the questions :

(a) Despite the saying that "a lawyer's profession is liar's profession", the writer (Gandhi) intended to———————————————

_____

(complete the sentence)

(b) Express in your own words the meaning of "put to the test".

(c) What was the temptation which Gandhi always resisted ?

(d) How could the client whose case was won have deceived Gandhi ?

(e) How much more did he charge for a case he won than for a case he lost ?

Q. 4. Read the following passage carefully :

A hundred years ago there was much less division of labour than there is today. One carpenter made the whole of a table–perhaps he made all the furniture necessary for a house. Most modern furniture is made in factories by machines. Nowadays a man working in a furniture factory may spend his whole time looking after one machine which can perform only a single operation, for example shaping the legs of a certain kind of table. A large number of men will work on other parts of the table, each performing a single operation. An experienced worker may understand all the operations which are necessary to make the table even if he himself does only one of them. But when a factory produces automobiles or electrical machines that require a great variety of parts, even this is not possible. No single individual can understand all the operations required to make the car or the machine. It is very likely that different factories will be responsible for making different parts. In modern life, therefore, everyone depends on everyone else's work in addition to his own. The danger of this is that a worker may begin to think that his own particular job is not important, that it does not matter very much how he does it. In the old days, when one, or even two or three people were responsible for the whole of a piece of work (for example, making a cart to be pulled by a horse), it was easy to say who was responsible if the cart was badly made or if it met with an accident because one part was not properly made. But nowadays, if a car has an accident because one of its parts was not properly made or finished, it is difficult to discover who was responsible for this.

Complete the following sentences using the information given in the passage :

(a) In a modern car factory even an experienced worker cannot———

_____.

(b) In a modern furniture factory, even though a man may understand several operations, he———————————————————.

(c) The different parts needed to make a car or electrical machine are often———————————————————————————.

(d) When there was less division of labour it was easier to say who was responsible if some———————————————————.

(e) Workers may feel that is does not matter how their own specific jobs are done because———————————————————.

Q. 5. Write a passage of about 250 words describing how you spent your time, while waiting at a big railway station for a train running about two hours late.

Q. 6. Write a passage of about 250 words discussing the qualities of a person whom you find interesting and inspiring. This could be a person who actually lived (or is living) or a character from fiction. Try and indicate why *you* find this person impressive.

---

The above sample of questions is intended to illustrate the nature of the difference—in complexity and difficulty – between the stages. A finished test in the new style for a particular level has not been presented. The marking scheme for Stage I requires some explanation. Each question is a block question, of many parts, carrying a weight of 4 marks. The parts of a question cover more or less the same area of linguistic content, and are of the same difficulty level. Since they represent minimum essentials, a clear indication of mastery of these points is needed. The basic rule would be that the student gets each and every part correct, i.e., earns all the points. However, to allow for possible bias in the selection of specific points, and other normal sources of fluctuation, this is relaxed to 75% of the points. So a student gets all marks for the question, or he gets zero. The student getting only about a third or even half of the parts correct may easily be dismissed as having only a superficial familiarity with the linguistic rule involved. The student getting about two-thirds correct is a borderline case, and chance appears to play a large part in his score. It is argued that, since there is some measure of repetition and similarity within each block question, a student who has the requisite mastery will get nearly all parts correct. It is unlikely that many borderline cases will actually be encountered. This claim is clearly subject to empirical verification.

A similar requirement of complete or consistent mastery underlies the passing level (effectively 32 marks) for the stage as a whole. A student who passes, clears at least 8 of the 10 block questions. In other words he shows near complete mastery (as specified above) of nearly all the various areas covered by the questions of stage I. The number of questions was set at 10

partly because of the convenient total (10 x 4=40) that this provides. Further, 10 questions of this nature would allow a fairly wide coverage of the areas that would be specified as constituting minimum requirements.

The purpose of the first stage being only to differentiate between those passing and those failing, the exact marks are no longer of any interest. All those who pass therefore get 40 marks (which is the base to which their marks in Stage II are added). The failures should logically get a mark of zero. However as a concession to convention (and the powerful sentiments that go with it) this failing mark can be arbitrarily set at 25 for every student.

The questions of Stage II are not very different from those that are found in many question papers today. A substantial difference in the order of difficulty between the two stages should be evident. For example all the vocabulary words in Question 5 (Stage I) and Question 1 (Stage II) are from the "new words" listed along with the lessons of the same Reader. In the former case, the context provided by the sentence makes the task of discrimination among the four alternatives easier, while this advantage is absent in the latter case. Another question (2) in Stage II requires the production of an original sentence to illustrate the meaning of the problem word or phrase. With some care, the scorer can make a distinction between a sentence that merely indicates what part of speech the word is, and one that effectively indicates its meaning. Two short comprehension passages are used which allow questions calling for some transformation of the exact words in the passage to be asked; the questions also call for the making of inferences using information from different sentences. Passages that are productive in terms of appropriate comprehension questions are extremely difficult to find. The use of two or three separate paragraphs rather than one lengthy passage (presumably with literary value) is regarded as more effective for the purposes of a second language test at the selected level.

Given a set of six questions each carrying 10 marks, a range of 60 marks for Stage II is available. The equivalence principle *does operate* here. However, because of the advanced level of all the questions the occurence of highly irregular profiles is unlikely. Few students are likely to do very well in some questions and very badly in others. A high degree of consistency in the marks (out of 10 ) for all the questions may be expected. This again is a proposition that can be studied empirically.

## VII. Summary and evaluation

In this paper an attempt has been made to assess the soundness of the existing system of examination, in relation to the actual purposes that are to be served. The fixed passmark is identified as a unique feature. The pass-fail decision is shown to be liable to serious eror. A scheme for making the pass-fail decision more rational and systematic has been suggested. The importance of specifying the minimum essentials in addition to the widely desiderated objectives, has been stressed. An examination model which uses two entirely separate stages is presented. It is held that the separation of the operations -- one to test against minimum requirements and the other to order the passing candidates into degrees of attainment -- is necessary. Some

sample questions to illustrate the nature of the two stages are included. The model requires that the questions in the two stages be substantially different in terms of difficulty level. Such an arrangement will require empirical tryout and establishment of difficulty levels. This practical and empirical step, in itself, would be a valuable addition to the existing paper-setting procedures.

Some assumptions about the hierarchical structure of language skills underly the model. These are open to empirical verification. These considerations inform the decisions made in applying the concept of gradation to lesson planning and the preparation of teaching materials. The specification of minimum requirements for passing (or "proceeding") is a major responsibility of teachers and curriculum builders. The submission is thus made that experimentation with the new approach to examining procedures proposed here, is of potential significance in areas outside the limited focus of examination reform.

*Bureau of Tests & Examinations,*
*CIEFL, Hyderabad–500007.*

## REFERENCES

Bloom B. S. (Ed.) (1956) *A Taxonomy of Educational Objectives. Handbook I : Cognitive Domain.* New York, Longmans Green.

Carrol J. B. (1961) Fundamental considerations in Testing for English Language Proficiency of Foreign Students. *Testing.* Centre for Applied Linguistics.

Cooper R. L. (1972) "Testing". *in* Allen H. B. & Campbell R. N. (Eds.) *Teaching English as a Second Language.* New York, McGraw Hill.

Cronbach L. J. (1970) "Validity of Educational Measures" in (Bracht. G. H. (Ed.) *Perspectives in Educational and Psychological Measurement.* Englewood Cliffs, Prentice Hall.

Furst E. J. (1958) *Constructing Evaluation Instruments.* New York, Longmans Green.

George H. V. (1961) "Statistics, School Leaving Examinations and Language Teaching". Bulletin of the C. I. E. No. 1.

Grieve D. W. 1964 *English Language Examining.* Lagos, African Universities Press

Mager R. F. 1962 *Preparing Instructional Objectives.* San Francisco, Fearon Publishers.

University Grants Commission (1973) *Examination Reform : A Plan of Action.* New Delhi, U. G. C.